



IFWO

RAW SEQUENCE LISTING

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:19

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

4 <110> APPLICANT: Bougueleret, Lydie
 5 Bairoch, Amos
 6 Niknejad, Anne
 8 <120> TITLE OF INVENTION: Engineered Human Kunitz-Type Protease
 9 Inhibitor
 11 <130> FILE REFERENCE: 54720-8015.US00
 13 <140> CURRENT APPLICATION NUMBER: US 10/807,204
 14 <141> CURRENT FILING DATE: 2004-03-22
 16 <150> PRIOR APPLICATION NUMBER: PCT/EP03/01629
 17 <151> PRIOR FILING DATE: 2003-02-18
 19 <150> PRIOR APPLICATION NUMBER: US 60/358,683
 20 <151> PRIOR FILING DATE: 2002-02-21
 22 <160> NUMBER OF SEQ ID NOS: 16
 24 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 26 <210> SEQ ID NO: 1
 27 <211> LENGTH: 131
 28 <212> TYPE: PRT
 29 <213> ORGANISM: Homo sapiens
 31 <220> FEATURE:
 32 <221> NAME/KEY: VARIANT
 33 <222> LOCATION: (1)...(131)
 34 <223> OTHER INFORMATION: eppin-like precursor
 36 <221> NAME/KEY: SIGNAL
 37 <222> LOCATION: (1)...(25)
 38 <223> OTHER INFORMATION: predicted by SignalP 2.0
 W--> 40 <221> PEPTIDE
 41 <222> LOCATION: (26)...(131)
 42 <223> OTHER INFORMATION: mature peptide
 W--> 44 <221> DOMAIN
 45 <222> LOCATION: (77)...(127)
 46 <223> OTHER INFORMATION: Kunitz domain predicted by pfscan
 W--> 48 <221> DISULFID
 49 <222> LOCATION: (33)...(61)
 50 <223> OTHER INFORMATION: predicted disulfide bond
 W--> 52 <221> DISULFID
 53 <222> LOCATION: (40)...(65)
 54 <223> OTHER INFORMATION: predicted disulfide bond
 W--> 56 <221> DISULFID
 57 <222> LOCATION: (48)...(60)
 58 <223> OTHER INFORMATION: predicted disulfide bond
 W--> 60 <221> DISULFID
 61 <222> LOCATION: (54)...(69)
 62 <223> OTHER INFORMATION: predicted disulfide bond



RAW SEQUENCE LISTING

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:19

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

```

W--> 64 <221> DISULFID
65 <222> LOCATION: (77)...(127)
66 <223> OTHER INFORMATION: predicted disulfide bond
W--> 68 <221> DISULFID
69 <222> LOCATION: (86)...(110)
70 <223> OTHER INFORMATION: predicted disulfide bond
W--> 72 <221> DISULFID
73 <222> LOCATION: (102)...(123)
74 <223> OTHER INFORMATION: predicted disulfide bond
W--> 76 <400> 1
77 Met Gly Leu Ser Gly Leu Leu Pro Ile Leu Val Pro Phe Ile Leu Leu
78 1 5 10 15
79 Gly Asp Ile Gln Glu Pro Gly His Ala Glu Gly Ile Leu Gly Lys Pro
80 20 25 30
81 Cys Pro Lys Ile Lys Val Glu Cys Glu Val Glu Glu Ile Asp Gln Cys
82 35 40 45
83 Thr Lys Pro Arg Asp Cys Pro Glu Asn Met Lys Cys Cys Pro Phe Ser
84 50 55 60
85 Cys Gly Lys Lys Cys Leu Asp Phe Arg Lys Asp Ile Cys Ser Met Pro
86 65 70 75 80
87 Gln Glu Ala Gly Pro Cys Leu Ala Ser Ile Pro His Trp Trp Tyr Asn
88 85 90 95
89 Lys Lys Thr Lys Ile Cys Ser Glu Phe Ile Tyr Gly Gly Cys Gln Gly
90 100 105 110
91 Asn Asn Asn Asn Phe Gln Thr Glu Ala Ile Cys Leu Val Thr Cys Lys
92 115 120 125
93 Lys Tyr His
94 130
96 <210> SEQ ID NO: 2
97 <211> LENGTH: 106
98 <212> TYPE: PRT
99 <213> ORGANISM: Homo sapiens
101 <220> FEATURE:
102 <221> NAME/KEY: VARIANT
103 <222> LOCATION: (1)...(106)
104 <223> OTHER INFORMATION: mature form
106 <221> NAME/KEY: DOMAIN
107 <222> LOCATION: (52)...(102)
108 <223> OTHER INFORMATION: Kunitz domain predicted by pfscan
W--> 110 <221> DISULFID
111 <222> LOCATION: (8)...(36)
112 <223> OTHER INFORMATION: predicted disulfide bond
W--> 114 <221> DISULFID
115 <222> LOCATION: (15)...(40)
116 <223> OTHER INFORMATION: predicted disulfide bond
W--> 118 <221> DISULFID
119 <222> LOCATION: (23)...(35)
120 <223> OTHER INFORMATION: predicted disulfide bond
W--> 122 <221> DISULFID

```

RAW SEQUENCE LISTING

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:19

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

```

123 <222> LOCATION: (29)...(44)
124 <223> OTHER INFORMATION: predicted disulfide bond
W--> 126 <221> DISULFID
127 <222> LOCATION: (52)...(102)
128 <223> OTHER INFORMATION: predicted disulfide bond
W--> 130 <221> DISULFID
131 <222> LOCATION: (61)...(85)
132 <223> OTHER INFORMATION: predicted disulfide bond
W--> 134 <221> DISULFID
135 <222> LOCATION: (77)...(98)
136 <223> OTHER INFORMATION: predicted disulfide bond
W--> 138 <400> 2
139 Glu Gly Ile Leu Gly Lys Pro Cys Pro Lys Ile Lys Val Glu Cys Glu
140 1 5 10 15
141 Val Glu Glu Ile Asp Gln Cys Thr Lys Pro Arg Asp Cys Pro Glu Asn
142 20 25 30
143 Met Lys Cys Cys Pro Phe Ser Cys Gly Lys Lys Cys Leu Asp Phe Arg
144 35 40 45
145 Lys Asp Ile Cys Ser Met Pro Gln Glu Ala Gly Pro Cys Leu Ala Ser
146 50 55 60
147 Ile Pro His Trp Trp Tyr Asn Lys Lys Thr Lys Ile Cys Ser Glu Phe
148 65 70 75 80
149 Ile Tyr Gly Gly Cys Gln Gly Asn Asn Asn Asn Phe Gln Thr Glu Ala
150 85 90 95
151 Ile Cys Leu Val Thr Cys Lys Lys Tyr His
152 100 105
154 <210> SEQ ID NO: 3
155 <211> LENGTH: 11201
156 <212> TYPE: DNA
157 <213> ORGANISM: Homo sapiens
159 <220> FEATURE:
160 <221> NAME/KEY: misc_feature
161 <222> LOCATION: (1)...(11201)
162 <223> OTHER INFORMATION: public sequence AL031663.2 (94000..105200)
164 <221> NAME/KEY: CDS
165 <222> LOCATION: (2043)...(2132)
W--> 167 <221> CDS
168 <222> LOCATION: (3358)...(3486)
W--> 170 <221> CDS
171 <222> LOCATION: (6120)...(6191)
W--> 173 <221> CDS
174 <222> LOCATION: (6193)...(6291)
W--> 176 <221> misc_feature
177 <222> LOCATION: (6291)...(6293)
178 <223> OTHER INFORMATION: potential stop codon
W--> 180 <221> 3'UTR
181 <222> LOCATION: (6294)...(7236)
182 <223> OTHER INFORMATION: partial
W--> 184 <221> polyA_signal

```

RAW SEQUENCE LISTING

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:19

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

185 <222> LOCATION: (7231)...(7236)

W--> 187 <400> 3

188	tttaaagaat	catgtaaaac	aggcttacaa	attttctttt	taaaaatatt	ttattttatt	60
189	ttacgttccg	ggatacatgt	gcaggacctg	caggtttggt	acataggtaa	gcctgtgcca	120
190	tgggtggttg	ctgcacctat	caaccattta	cctgggtatt	aagccccgca	tgcattagct	180
191	atztatcctg	atgctttccc	tccccctgcc	ccctcaacag	tccacgtggt	ctcattgttc	240
192	atctcccact	tgcaagtgag	aacatgtggt	atttactttt	gttcctgctg	tactttgctg	300
193	aggataacgg	cttcacgctt	catccatgtc	cctgcaaagg	acgtgatctc	atttcttttt	360
194	atggctgcat	agtattccat	ggcatatatg	taccacattt	tctttatcca	gcctatcggt	420
195	gatggatatt	tggtttgatt	ccatgttttt	gctattgtaa	atagtgtctg	agtgaacata	480
196	cacatgcatg	tatctttata	aaagaatgat	ttatattcct	ttgagtgtat	accagtaat	540
197	gggattgctg	ggtcaaattg	tatttctggt	tctagatctt	tcaggaattg	gcacactctc	600
198	ttccacaatg	gttgaaccaa	tttacattcc	caacagtgtg	aaagcattca	tgtttctcca	660
199	cagctttacc	agcatctggt	gtttctggac	tttttaataa	tcaccattct	gactggcatg	720
200	agatggatgc	tcattgtggt	tttgatttgc	atttctctga	tgatcactgg	tgttgagctt	780
201	tttttcatat	gtttgttggc	cgcataaatg	tcttcttttg	agaagtgtct	gttcattgtc	840
202	tttgcccact	tttcaatggg	gttggttggg	tttttcttg	taaatttggt	taagttcctt	900
203	gtagattctg	gatatcagac	cacaaatttt	cttaactctg	ctccagtgat	gggaaagcta	960
204	agctatttgt	ccactgattc	ccaacgcccc	ctggatgcca	gggatactaa	atccctcgct	1020
205	ctttcaaact	gtttatactg	caagctatac	aagctccctg	ggggccagag	aaaaccacac	1080
206	gcagaaaatt	agggactgat	gctagaggta	ggaggctgtc	aatgtgcatt	ggctattatt	1140
207	cggccataaa	aaggaatgaa	cttctcactc	accactgatg	acatgccttg	ctgcacacac	1200
208	accctgctgc	caccctgcag	aaatgcttec	attaccacac	gtcctttgcc	agatggaact	1260
209	gatgcccagg	taactggctc	ctcacctcct	tcgggcttat	accaaggata	taaaacttgg	1320
210	tacagatgct	caagcgttaa	tgcttcaggg	tgtgactttt	aatccaatgc	tgtagctgtg	1380
211	acaacaggga	aaaacggaga	acagagttgg	agaagtccca	aagtaatgaa	agaagttatt	1440
212	tactgttgca	aaatcaaacg	acttaaagga	taaatgtaac	tgatctttac	atacgaaaga	1500
213	agagagaggg	aagtgaaggg	cggggagaaa	gagactcaga	cagacagaga	ctccctgaga	1560
214	aaaagacagg	cacagagtaa	aagacatata	gatgcacaga	aaagagaaga	aacagagtga	1620
215	gaaaaataaa	acataagaat	agacatgtca	ggagaagagt	gaaattagaa	agaaggaaga	1680
216	gagctagcat	actgttaacc	tgcaaaactc	ccctgggaaa	tggaagactc	tggaagttaa	1740
217	cctccccttc	tgctaaggga	gcctggaaaag	cccaagacaa	catgtgccct	ctttccctgg	1800
218	tagaccagaa	agcctcagat	tgctgaggat	taggagccac	ctatccccc	gtgtaggtac	1860
219	caggtagtcc	cactgatggc	ctcagggatc	tggtgttcc	gatcggggag	tctggctggc	1920
220	ttgtctccag	ggctgtctct	taaggtgctg	ggccacaccg	gggcaggcaa	agtgcagatg	1980
221	taggagctgc	tgaggagcag	ttcttagttt	tggtgccatc	aaccaggcca	gtacttccta	2040
222	aa atg gga	ctc tca gga	ctt ctg cca	atc ctg gta	cca ttc atc	ctt	2087
223	Met Gly	Leu Ser Gly	Leu Leu Pro	Ile Leu Val	Pro Phe	Ile Leu	
224	1	5	10	15			
226	ttg ggg gac	atc cag gaa	cct ggg cac	gct gaa ggc	atc ctt ggc		2132
227	Leu Gly Asp	Ile Gln Glu	Pro Gly His	Ala Glu Gly	Ile Leu Gly		
228	20	25	30				
230	agtaagtatt	gggaactccg	tccttgcaat	ggggaaagga	aactcagtgg	ggaggaagga	2192
231	acaaagaata	attttgcagc	tgagaaatta	ttttctccac	ctctgcctgg	aattccaggt	2252
232	gggaaagagg	ggagctattt	gaagtctctt	ggtaagagac	tcccttggtg	aatacaggca	2312
233	tccaaatctg	tggcaagtga	ctccttgctg	ccattccaaa	gagttcttcc	tggaggtgcc	2372
234	tggacctaga	gctcatgaaa	tgaactaaga	aggagtagta	agcccaagga	agtctccac	2432
235	aagaaaaaca	ggctgctccc	cacaacacca	cctgtgtcaa	cttccaaatg	ttctcattgc	2492
236	agcccccaca	cccataccca	cagaggaccc	tggtctttgg	gttaaagaat	ggtctccaga	2552

RAW SEQUENCE LISTING

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:19

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

```

237 agtcccttgg aatgctgaaa tgtgaagact cacaattagc agctaactgt actgggcact 2612
238 gaaaattagt ggttgaaagt gagcactcag gagccacaca ttgtagggtc aaacctggct 2672
239 ccaccctga gtagctgtgt gatcttgggc aaattcacat aatctcctta tgcttccaat 2732
240 tcttctgttt acaaaatgtg aatattaaga gtacttacc c atagatttgt aatgaggatt 2792
241 aaggagaaaa ggcagctcac agtaagtgtc tcatccatat tactcattgt tacgtttacat 2852
242 tgccccaggc agtttgacga tggcttgctc aactgtgttg ctggttagtc ttcaaagcaa 2912
243 ccctatgggc aaatgaggaa atcaaggctc agagagagtg cctggcacac caagcgtctc 2972
244 acagaattca aaagcccagc tatctgatgc cagaaatgtt aatctcaatc atcgagagcc 3032
245 acagcacctc ctggggggagc aaccaagcac agcaaggctg agtgaccaga acagtcttag 3092
246 aggaatgcgt ctttctctca ggaatctcat agatggtgcg ggtagggggg aggggttggt 3152
247 ttaggcaccc gatggtgagt tgttatgggg ttggggatgt gagcagtgtt gtgaaattcc 3212
248 tgagccccct gagggccact accttctggg gataggtgaa gtggaaacaa agttctctaa 3272
249 ctctggaggg gtgccctggt ctgggggagg gggagcactg gctcctttaa ggccttgagg 3332
250 gaatatatct tccccctcct tagag ccg tgt ccc aaa atc aaa gtg gaa tgc 3384
251                                     Pro Cys Pro Lys Ile Lys Val Glu Cys
252                                     35
254 gaa gtg gaa gaa ata gac cag tgt acc aaa ccc aga gat tgc cca gaa 3432
255 Glu Val Glu Glu Ile Asp Gln Cys Thr Lys Pro Arg Asp Cys Pro Glu
256 40 45 50 55
258 aac atg aag tgt tgc ccg ttc agc cgt gga aag aaa tgt tta gac ttc 3480
259 Asn Met Lys Cys Cys Pro Phe Ser Arg Gly Lys Lys Cys Leu Asp Phe
260 60 65 70
262 aga aag gtaactcaga tgcttcttaa attaccacgt gccctcacct cctatctcca 3536
263 Arg Lys
266 cctgcactgc actacctctg tctcctagtt cactgatggc tggctctctga ccaagagttt 3596
267 aggtacatct tctacttgac acaaactctgc cagtcctgga ctggggagac atggatttta 3656
268 ttctatttct tctctatatt tggtaagtga tgtgaacaat gtagttttcc tttctgggcc 3716
269 tcagtttccc tttccataag acaggtgtgg ataatggacc acatgggtctc ttggctactc 3776
270 ttcagtgtat ctcgagatgc caagtagtca ccaaagtctc atcaaagcc aacatttgtt 3836
271 caggttatat gatagctata agagacaggt atataatctg tatttccttg cttttttctc 3896
272 taggcctctc tgtctacata aatccaggat gccttccctg accaccttag accgcctaca 3956
273 ctctttcttc tctgcccata ttattgagct ctcaaagaaga tcatcaagga ccattcgttc 4016
274 attctattca ttcaacaaat attgattggt gccctgcaga cccatgatgg gttccacttg 4076
275 ctcatgttgg ccaggaaggt cagaaaagtt tttctgggaa agagatccat gatttaagat 4136
276 ctgaaagatg agtacttaga ccaaaaatgg gaagaaatag cctttcaggc cagataatag 4196
277 gatgtgcaaa ggctctgtgg caggaggcag gatgttcagg tctagaaagg ctagaaagca 4256
278 gagagtgagc aagagacaag aaacaggaaa ctggtagtgt atattcaggg caacttgaat 4316
279 ctgtattccc aagtggccat cctcaagttt tgggggtcaaa taaatgccct acttaatgat 4376
280 attttctgaa tcttggtact ttaggttgac atacctttca agacaataac taggttttca 4436
281 gtgctttgat aaaacctcaa atgtgtggcc ttccagccag gcactctgtg tttcatattg 4496
282 ttaagcatt tgttgcatcc ctatggattg cagctctcta taggttcgct gttgctgaac 4556
283 caagaaataa ggcatacaca gatttctaga gaaattgtaa ttctataaca gattttcagg 4616
284 agatgaggaa ggctgggtatc tttcagagta caaagtgtat ttggaattga aagaatttct 4676
285 ttgggttaaa gtttgtcact gacttgtgtt cctgaaccat gaagcatgaa tatgtgggct 4736
286 gagaaatagt ttctcttgat aaataaacia gtaacaaatt ttaaaaagag aggaagatga 4796
287 agctgggcta gttgggggaa ggagagctgg taagggatct ggtcttgatc ctaacagcaa 4856
288 tggattaacc tatgttcaaa tcccttacca tctgtgtggc ctcaagaaag ttatttagcc 4916
289 tctctaaacc ttagttttct catctgcata atgagaacaa tgagaggacc tactttacaa 4976
290 gattaaatgt taatagatgt caactactaa gaacactaat caatgctata attgttggtt 5036

```

VERIFICATION SUMMARY

DATE: 07/22/2004

PATENT APPLICATION: US/10/807,204

TIME: 10:25:20

Input Set : A:\54720-8015.US00-SEQLIST.TXT

Output Set: N:\CRF4\07222004\J807204.raw

L:40 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:44 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:48 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:52 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:56 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:60 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:64 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:68 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:72 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:76 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:110 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:114 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:118 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:122 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:126 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:130 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:134 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:138 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:167 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:170 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:173 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:176 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:180 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:184 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:187 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:469 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:8
L:472 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:8
L:476 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:8
L:480 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:8
L:483 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:8
L:584 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:11